

SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: Koopman, Peter  
Goodfellow, Peter
- (ii) TITLE OF THE INVENTION: SOX-9 GENE AND PROTEIN AND  
USE IN THE REGENERATION OF BONE OR CARTILAGE
- (iii) NUMBER OF SEQUENCES: 21
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Scully, Scott, Murphy & Presser
  - (B) STREET: 400 Garden City Plaza
  - (C) CITY: Garden City
  - (D) STATE: NY
  - (E) COUNTRY: U.S.A.
  - (F) ZIP: 11530
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Diskette
  - (B) COMPUTER: IBM Compatible
  - (C) OPERATING SYSTEM: DOS
  - (D) SOFTWARE: FastSEQ Version 1.5
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: 08/860,635
  - (B) FILING DATE: 29-MAY-1997
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: AU PM9714
  - (B) FILING DATE: 29-NOV-1994
  - (A) APPLICATION NUMBER: AU PM9835
  - (B) FILING DATE: 05-DEC-1994
  - (A) APPLICATION NUMBER: PCT/AU95/00799
  - (B) FILING DATE: 29-NOV-1995
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: DiGiglio, Frank S.
  - (B) REGISTRATION NUMBER: 31,346
  - (C) REFERENCE/DOCKET NUMBER: 10981
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 516-742-4343
  - (B) TELEFAX: 516-742-4366
  - (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AATTAAA

7

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CCAAAGTCCT AAAGGTGGG

19

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TTTCAGGCAA ATAAGGCAG

19

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TGGCAATCTA ACAGATGAGA

20

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TCNCAAATGT CATATATCCA

20

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGTCCAGATT GACTGGAACCA CA

22

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCAATAAGAT ACTAATATGT AGAG

24

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GTCAGCAGAA ATCCTAAAGG

20

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
D (D) TOPOLOGY: linear

D (ii) MOLECULE TYPE: cDNA

D (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

D GACTAATGCC GATGGTTAAG

20

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
D (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CGCCTCGAGG TGGCTTATCG

20

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
D (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ATCATAACACA TACGATTTAG GTGAC

25

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GAGGAAGTCG GTGAAGAAC

19

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TCGCTCATGC CGGAGGAGGA G

21

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCAATCCCAG GGCCCACCGA C

21

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TTGGAGATGA CGTCGACTGC TC

22

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GCAGCGACGT CATCTCCAAC

20

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GCTGCTTGGA CATCCACACG T

21

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2249 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

AGTTTCAGTC	CAGGAAC	TTT	TCTTGCAAG	AGAGACGAGG	TGCAAGTGGC	50
CCC GGTTTCG	TTCTCTGTT	TCCCTCCCTC	CTCCTCCGCT	CCGACTCGCC		100
TTCCCCGGGT	TTAGAGCCGG	CAGCTGAGAC	CCGCCACCCA	GCGCCTCTGC		150
TAAGTGCCCG	CCGCCGCAGC	CCGGTGACGC	GCCAACCTCC	CCGGGAGCCG		200
TTCGCTCGGC	GTCCGGGTCC	GGGCAGCTGA	GGGAAGAGGA	GCCCCAGCCG		250
CCGCGGCTTC	TCGCCTTCC	CGGCCACCCG	CCCCCTGCC	CGGGCTCGCG		300
TATGAATCTC	CTGGACCCCT	TCATGAAGAT	GACCGACGAG	CAGGAGAAGG		350
GCCTGTCTGG	CGCCCCCAGC	CCCACCATGT	CGGAGGACTC	GGCTGGTTCG		400
CCCTGTCCCT	CGGGCTCCGG	CTCGGACACG	GAGAACACCC	GGCCCCAGGA		450
GAACACCTTC	CCCAAGGGCG	AGCCGGATCT	GAAGAAGGAG	AGCGAGGAAG		500
ATAAGTTCCC	CGTGTGCATC	CGCGAGGC	TCAGCCAGGT	GCTGAAGGGC		550
TACGACTGGA	CGCTGGTGCC	CATGCCCGTG	CGCGTCAACG	GCTCCAGCAA		600
GAACAAGCCA	CACGTCAAGC	GACCCATGAA	CGCCTTCATG	GTGTGGCGC		650
AGGCTGCGCG	CAGGAAGCTG	GCAGACCAGT	ACCCGCATCT	GCACAAACGCG		700
GAGCTCAGCA	AGACTCTGGG	CAAGCTCTGG	AGGCTGCTGA	ACGAGAGCGA		750
GAAGAGACCC	TTCGTGGAGG	AGGCGGAGCG	GCTGCGCGTG	CAGCACAAAGA		800
AAGACCACCC	CGATTACAAG	TACCAGCCCC	GGCGGAGGAA	GTCGGTGAAG		850
AACGGACAAG	CGGAGGCCGA	AGAGGCCACG	GAACAGACTC	ACATCTCTCC		900
TAATGCTATC	TTCAAGGCGC	TGCAAGCCGA	CTCCCCACAT	TCCTCCTCCG		950
GCATGAGTGA	GGTGCACTCC	CCGGGCGAGC	ACTCTGGCA	ATCTCAGGGT		1000
CCGCCGACCC	CACCCACAC	TCCCAAAACC	GACGTGCAAG	CTGGCAAAGT		1050
TGATCTGAAG	CGAGAGGGGC	GCCCTCTGGC	AGAGGGGGC	AGACAGCCCC		1100
CCATCGACTT	CCGCGACGTG	GACATCGGTG	AACTGAGCAG	CGACGTCATC		1150
TCCAACATTG	AGACCTTCGA	CGTCAATGAG	TTTGACCAAT	ACTTGCCACC		1200
CAACGGCCAC	CCAGGGGTTTC	CGGCCACCCA	CGGCCAGGTC	ACCTACACTG		1250
GCAGTTACGG	CATCAGCAGC	ACCGCACCCA	CCCCTGCGAC	CGCGGGCCAC		1300
GTGTGGATGT	CGAAGCAGCA	GGCGCCGCC	CCTCCTCCGC	AGCAGCCTCC		1350
GCAGGGCCCCG	CAAGCCCCAC	AGGCGCCTCC	GCAGCAGCAA	GCACCCCCGC		1400

AGCAGCCGCA	GGCACCCCCAG	CAGCAGCAGG	CACACACGCT	CACCACGCTG	1450
AGCAGCGAGC	CAGGCCAGTC	CCAGCGAACG	CACATCAAGA	CGGAGCAGCT	1500
GAGCCCCAGC	CACTACAGGG	AGCAGCAGCA	GCACACTCCCCG	CAACAGATCT	1550
CCTACAGCCC	CTTCAACCTT	CCTCACTACA	GGCCCTCCTA	CCCGCCCCATC	1600
ACCCGTTCGG	AATACGACTA	CGCTGACCAT	CAGAACTCCG	GCTCCTACTA	1650
CAGTCACGCA	GCCGGCCAGG	GCTCAGGGCT	CTACTCCACC	TTCACTTACA	1700
TGAACCCCGC	GCAGCGCCCC	ATGTACACCC	CCATCGGTGA	CACCTCCGGG	1750
GTCCCTTCCA	TCCCCGAGAC	CCACAGCCCG	CAGGACTGGG	AACAACCAGT	1800
CTACACACAG	GTCACCAGAC	CCTGAGAAGA	GAAAAGCTAT	GGTGACAGAG	1850
CTGATCTTTT	TTTTTTTTTT	TTTTTAAAGA	AGAAAAGAAA	GAAACGAAAA	1900
AGAAAAAGCT	GAAGGAAATC	AAGAACCAAT	TGAAATTCTT	TTGGACACTT	1950
TTTTTTTGT	CCTTCGTTA	ATTTTAAAAA	GACATGTAAA	GGAAGGTAAC	2000
GATTGCTGGG	CATTCCAGGA	GAGAGACTTT	AAGACTTTGT	CTGAGCTCAT	2050
GACAACATAT	TGCAAATGGC	CGGGCCACTC	GTGGCCAGAC	GGACAGCACT	2100
CCTGGCCAGA	TGGACCCACC	AGTATCAGCG	AGGAGGGGCT	TGTCTCCTTC	2150
AGAGTTAACAA	TGGAGGACGA	TTGGAGAACATC	TCCCTGCCTG	TTTGGACTTT	2200
GTAATTATT	TTTAGCCGTA	ATTAAGAAA	AAAAAAAGTCC	AAAAAAAAAA	2249

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:

  - (A) LENGTH: 507 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Met	Asn	Leu	Leu	Asp	Pro	Phe	Met	Lys	Met	Thr	Asp	Glu	Gln	Glu	Lys
1									10						15
Gly	Leu	Ser	Gly	Ala	Pro	Ser	Pro	Thr	Met	Ser	Glu	Asp	Ser	Ala	Gly
			20					25					30		
Ser	Pro	Cys	Pro	Ser	Gly	Ser	Gly	Ser	Asp	Thr	Glu	Asn	Thr	Arg	Pro
								35				40		45	

Gln Glu Asn Thr Phe Pro Lys Gly Glu Pro Asp Leu Lys Lys Glu Ser  
50 55 60

Glu Glu Asp Lys Phe Pro Val Cys Ile Arg Glu Ala Val Ser Gln Val  
65 70 75 80

Leu Lys Gly Tyr Asp Trp Thr Leu Val Pro Met Pro Val Arg Val Asn  
85 90 95

Gly Ser Ser Lys Asn Lys Pro His Val Lys Arg Pro Met Asn Ala Phe  
100 105 110

Met Val Trp Ala Gln Ala Ala Arg Arg Lys Leu Ala Asp Gln Tyr Pro  
115 120 125

His Leu His Asn Ala Glu Leu Ser Lys Thr Leu Gly Lys Leu Trp Arg  
130 135 140

Leu Leu Asn Glu Ser Glu Lys Arg Pro Phe Val Glu Glu Ala Glu Arg  
145 150 155 160

Leu Arg Val Gln His Lys Lys Asp His Pro Asp Tyr Lys Tyr Gln Pro  
165 170 175

Arg Arg Arg Lys Ser Val Lys Asn Gly Gln Ala Glu Ala Glu Glu Ala  
180 185 190

Thr Glu Gln Thr His Ile Ser Pro Asn Ala Ile Phe Lys Ala Leu Gln  
195 200 205

Ala Asp Ser Pro His Ser Ser Ser Gly Met Ser Glu Val His Ser Pro  
210 215 220

Gly Glu His Ser Gly Gln Ser Gln Gly Pro Pro Thr Pro Pro Thr Thr  
225 230 235 240

Pro Lys Thr Asp Val Gln Ala Gly Lys Val Asp Leu Lys Arg Glu Gly  
245 250 255

Arg Pro Leu Ala Glu Gly Gly Arg Gln Pro Pro Ile Asp Phe Arg Asp  
260 265 270

Val Asp Ile Gly Glu Leu Ser Ser Asp Val Ile Ser Asn Ile Glu Thr  
275 280 285

Phe Asp Val Asn Glu Phe Asp Gln Tyr Leu Pro Pro Asn Gly His Pro  
290 295 300

Gly Val Pro Ala Thr His Gly Gln Val Thr Tyr Thr Gly Ser Tyr Gly  
305 310 315 320

Ile Ser Ser Thr Ala Pro Thr Pro Ala Thr Ala Gly His Val Trp Met  
325 330 335

Ser Lys Gln Gln Ala Pro Pro Pro Pro Pro Gln Gln Pro Pro Gln Ala  
340 345 350

Pro Gln Ala Pro Gln Ala Pro Pro Gln Gln Gln Ala Pro Pro Gln Gln  
 355 360 365  
 Pro Gln Ala Pro Gln Gln Gln Ala His Thr Leu Thr Thr Leu Ser  
 370 375 380  
 Ser Glu Pro Gly Gln Ser Gln Arg Thr His Ile Lys Thr Glu Gln Leu  
 385 390 395 400  
 Ser Pro Ser His Tyr Arg Glu Gln Gln Gln His Ser Pro Gln Gln Ile  
 405 410 415  
 Ser Tyr Ser Pro Phe Asn Leu Pro His Tyr Arg Pro Ser Tyr Pro Pro  
 420 425 430  
 Ile Thr Arg Ser Glu Tyr Asp Tyr Ala Asp His Gln Asn Ser Gly Ser  
 435 440 445  
 Tyr Tyr Ser His Ala Ala Gly Gln Gly Ser Gly Leu Tyr Ser Thr Phe  
 450 455 460  
 Thr Tyr Met Asn Pro Ala Gln Arg Pro Met Tyr Thr Pro Ile Gly Asp  
 465 470 475 480  
 Thr Ser Gly Val Pro Ser Ile Pro Gln Thr His Ser Pro Gln Asp Trp  
 485 490 495  
 Glu Gln Pro Val Tyr Thr Gln Val Thr Arg Pro  
 500 505

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3923 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CGGAGCTCGA AACTGACTGG AAACCTTCAGT GGCGCGGAGA CTCGCCAGTT TCAACCCCGG	60
AAACTTTCT TTGCAGGAGG AGAACAGAGAAG GGGTGCAAGC GCCCCCCACTT TTGCTCTTTT	120
TCCTCCCCTC CTCCTCCTCT CCAATTGCC TCCCCCCACT TGGAGCGGGC AGCTGTGAAC	180
TGGCCACCCC GCGCCTTCCT AAGTGCTCGC CGCGGTAGCC GGCCGACGCG CCAGCTTCCC	240
CGGGAGCCGC TTGCTCCGCA TCCGGGCAGC CGAGGGGAGA GGAGCCCGCG CCTCGAGTCC	300
CCGAGCCGCC GCGGCTTCTC GCCTTCCCCG GCCACCAGCC CCCTGCCCG GGCCCGCGTA	360
TGAATCTCCT GGACCCCTTC ATGAAGATGA CCGACGAGCA GGAGAAGGGC CTGTCCGGCG	420

CCCCCAGCCC CACCATGTCC GAGGACTCCG CGGGCTCGCC CTGCCCGTCG GGCTCCGGCT 480  
CGGACACCGA GAACACCGCG CCCCAGGAGA ACACGTTCCC CAAGGGCGAG CCCGATCTGA 540  
AGAAGGAGAG CGAGGAGGAC AAGTTCCCCG TGTGCATCCG CGAGGCGGTC AGCCAGGTGC 600  
TCAAAGGCTA CGACTGGACG CTGGTGCCCA TGCCGGTGCG CGTCAACGGC TCCAGCAAGA 660  
ACAAGCCGCA CGTCAAGCGG CCCATGAACG CCTTCATGGT GTGGGCGCAG GCGGCGCGCA 720  
GGAAGCTCGC GGACCAGTAC CCGCACTTGC ACAACGCCGA GCTCAGCAAG ACGCTGGGCA 780  
AGCTCTGGAG ACTTCTGAAC GAGAGCGAGA AGCGGCCCTT CGTGGAGGAG GCGGAGCGGC 840  
TGC CGGTGCA GCACAAGAAG GACCACCCGG ATTACAAGTA CCAGCCGGCG CGGAGGAAGT 900  
CGGTGAAGAA CGGGCAGGCG GAGGCAGAGG AGGCCACGGA GCAGACGCAC ATCTCCCCA 960  
ACGCCATCTT CAAGGCGCTG CAGGCCGACT CGCCACACTC CTCCTCCGGC ATGAGCGAGG 1020  
TGC ACTCCCC CGGCGAGCAC TCGGGGCAAT CCCAGGGCCC ACCGACCCCA CCCACCACCC 1080  
CAAAAACCGA CGTGCAGCCG GGCAAGGCTG ACCTGAAGCG AGAGGGCGC CCCTTGCCAG 1140  
AGGGGGGCAG ACAGCCCCCT ATCGACTTCC GCGACGTGGA CATCGGCGAG CTGAGCAGCG 1200  
ACGTCATCTC CAACATCGAG ACCTTCGATG TCAACGAGTT TGACCAGTAC CTGCCGCCA 1260  
ACGGCCACCC GGGGGTGCCG GCCACGCACG GCCAGGTAC CTACACGGC AGCTACGGCA 1320  
TCAGCAGCAC CGCGGCCACC CGGGCGAGCG CGGGCCACGT GTGGATGTCC AAGCAGCAGG 1380  
CGCCGCCGCC ACCCCCCGAG CAGCCCCCAC AGGCCCCGCC GGCCCCGCCAG GCGCCCCCGC 1440  
AGCCGCAGGC GGCGCCCCCA CAGCAGCCGG CGGCACCCCC GCAGCAGCCA CAGGCGCACA 1500  
CGCTGACCAC GCTGAGCAGC GAGCCGGGCC AGTCCCAGCG AACGCACATC AAGACGGAGC 1560  
AGCTGAGCCC CAGCCACTAC AGCGAGCAGC AGCAGCACTC GCCCCAACAG ATCGCCTACA 1620  
GCCCTTCAA CCTCCCCACAC TACAGCCCCCT CCTACCCGCC CATCACCCGC TCACAGTACG 1680  
ACTACACCGA CCACCAAGAAC TCCAGCTCCT ACTACAGCCA CGCGGCAGGC CAGGGCACCG 1740  
GCCTCTACTC CACCTTCACC TACATGAACC CCGCTCAGCG CCCCATGTAC ACCCCCCATCG 1800  
CCGACACCTC TGGGGTCCCT TCCATCCCGC AGACCCACAG CCCCCAGCAC TGGGAACAAC 1860  
CCGTCTACAC ACAGCTCACT CGACCTTGAG GAGGCCTCCC ACGAAGGGCG ACGATGGCCG 1920  
AGATGATCCT AAAAATAACC GAAGAAAGAG' AGGACCAACC AGAATTCCCT TTGGACATT 1980  
GTGTTTTTT GTTTTTTAT TTTGTTTGT TTTTTCTTCT TCTTCTTCTT CCTTAAAGAC 2040  
ATTAAAGCTA AAGGCAACTC GTACCCAAAT TTCCAAGACA CAAACATGAC CTATCCAAGC 2100

GCATTACCCA CTTGTGGCCA ATCAGTGGCC AGGCCAACCT TGGCTAAATG GAGCAGCGAA 2160  
ATCAACGAGA AACTGGACTT TTTAAACCCCT CTTCAGAGCA AGCGTGGAGG ATGATGGAGA 2220  
ATCGTGTGAT CAGTGTGCTA AATCTCTCTG CCTGTTGGA CTTTGTATT ATTTTTTAG 2280  
CAGTAATTAA AGAAAAAAAGT CCTCTGTGAG GAATATTCTC TATTTAAAT ATTTTTAGTA 2340  
TGTACTGTGT ATGATTTCATT ACCATTGAGA GGGGATTAT ACATATTTT AGATAAAATT 2400  
AAATGCTCTT ATTTTCCAA CAGCTAAACT ACTCTTAGTT GAACAGTGTG CCCTAGCTTT 2460  
TCTTGCAACC AGAGTATT TGACAGATT TGCTTCTCT TACAAAAGA AAAAAGAAT 2520  
CCTGTTGTAT TAACATTAA AACAGAATT GTGTTATGTG ATCAGTTTG GGGGTTAACT 2580  
TTGCTTAATT CCTCAGGCTT TGCGATTAA GGAGGAGCTG CCTTAAAAAA AAATAAAGGC 2640  
CTTATTTGC AATTATGGGA GTAAACAATA GTCTAGAGAA GCATTTGGTA AGCTTTATGA 2700  
TATATATATT TTTTAAAGAA GAGAAAAACA CCTTGAGCCT TAAAACGGTG CTGCTGGAA 2760  
ACATTTGCAC TCTTTAGTG CATTCCCTCC TGCCTTGCT TGTCACTGC AGTCTTAAGA 2820  
AAGAGGTAAA AGGCAAGCAA AGGAGATGAA ATCTGTTCTG GGAATGTTTC AGCAGCCAAT 2880  
AAGTCCCCGA GCACACTGCC CCCGGTTGCC TGCCTGGGCC CCATGTGGAA GGCAGATGCC 2940  
TGCTCGCTCT GTCACCTGTG CCTCTCAGAA CACCAGCAGT TAACCTTCAA GACATTCCAC 3000  
TTGCTAAAAT TATTTATT TGAAAGGAGAG GTTTAATTA AAACAAAAAA AAATTCTTT 3060  
TTTTTTTTTT TTTTCCAATT TTACCTTCTT TAAAATAGGT TGTTGGAGCT TTCCTCAAAG 3120  
GGTATGGTCA TCTGTTGTTA AATTATGTTC TTAACTGTAA CCAGTTTTT TTTATTTATC 3180  
TCTTTAATCT TTTTATTAT TAAAAGCAAG TTTCTTGTA TTCCTCACCC TAGATTGTA 3240  
TAAATGCCTT TTTGTCCATC CCTTTTTCT TTGTTGTTTG TGTTGAAAAC AAACCTGGAAA 3300  
CTTGTTCCTT TTTTGATATA AATGAGAGAT TGCAAATGTA GTGTATCACT GAGTCATTG 3360  
CAGTGTTC TGCCACAGAC CTTGGGCTG CCTTATATTG TGTGTGTGTG TGGGTGTGTG 3420  
TGTGTTCGA CACAAAAACA ATGCAAGCAT GTGTCATCCA TATTTCTCTA CATCTCTCT 3480  
TGGAGTGAGG GAGGCTACCT GGAGGGGATC AGCCCAGTGA CAGACCTAA TCTTAATTAC 3540  
TGCTGTGGCT AGAGAGTTG AGGATTGCTT TTTAAAAAG ACAGCAAAC TTTTTTTTA 3600  
TTTAAAAAA GATATATTAA CAGTTTAGA AGTCAGTAGA ATAAAATCTT AAAGCACTCA 3660  
TAATATGGCA TCCTTCAATT TCTGTATAAA AGCAGATCTT TTTAAAAAG ATACTTCTGT 3720  
AACTTAAGAA ACCTGGCATT TAAATCATAT TTTGTCTTA GGTAAAAGCT TTGGTTGTG 3780

TTCGTGTTT	GTTTGTTCA	CTTGTTCCC	TCCCAGCCCC	AAACCTTTG	TTCTCTCCGT	3840
GAAACTTACC	TTTCCCTTT	TCTTCTCTT	TTTTTTTTG	TATATTATTG	TTTACAATAA	3900
ATATACATTG	CATTAAAAAG	AAA				3923

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 509 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Met	Asn	Leu	Leu	Asp	Pro	Phe	Met	Lys	Met	Thr	Asp	Glu	Gln	Glu	Lys
1				5			10		15						
Gly	Leu	Ser	Gly	Ala	Pro	Ser	Pro	Thr	Met	Ser	Glu	Asp	Ser	Ala	Gly
20							25							30	
Ser	Pro	Cys	Pro	Ser	Gly	Ser	Gly	Ser	Asp	Thr	Glu	Asn	Thr	Arg	Pro
35							40					45			
Gln	Glu	Asn	Thr	Phe	Pro	Lys	Gly	Glu	Pro	Asp	Leu	Lys	Lys	Glu	Ser
50						55					60				
Glu	Glu	Asp	Lys	Phe	Pro	Val	Cys	Ile	Arg	Glu	Ala	Val	Ser	Gln	Val
65						70			75				80		
Leu	Lys	Gly	Tyr	Asp	Trp	Thr	Leu	Val	Pro	Met	Pro	Val	Arg	Val	Asn
85							90					95			
Gly	Ser	Ser	Lys	Asn	Lys	Pro	His	Val	Lys	Arg	Pro	Met	Asn	Ala	Phe
100							105					110			
Met	Val	Trp	Ala	Gln	Ala	Ala	Arg	Arg	Lys	Leu	Ala	Asp	Gln	Tyr	Pro
115							120					125			
His	Leu	His	Asn	Ala	Glu	Leu	Ser	Lys	Thr	Leu	Gly	Lys	Leu	Trp	Arg
130						135					140				
Leu	Leu	Asn	Glu	Ser	Glu	Lys	Arg	Pro	Phe	Val	Glu	Glu	Ala	Glu	Arg
145						150					155			160	
Leu	Arg	Val	Gln	His	Lys	Lys	Asp	His	Pro	Asp	Tyr	Lys	Tyr	Gln	Pro
							165			170			175		
Arg	Arg	Arg	Lys	Ser	Val	Lys	Asn	Gly	Gln	Ala	Glu	Ala	Glu	Glu	Ala
							180			185			190		
Thr	Glu	Gln	Thr	His	Ile	Ser	Pro	Asn	Ala	Ile	Phe	Lys	Ala	Leu	Gln
							195			200			205		
Ala	Asp	Ser	Pro	His	Ser	Ser	Ser	Gly	Met	Ser	Glu	Val	His	Ser	Pro
							210			215			220		
Gly	Glu	His	Ser	Gly	Gln	Ser	Gln	Gly	Pro	Pro	Thr	Pro	Pro	Thr	Thr
225							230				235			240	
Pro	Lys	Thr	Asp	Val	Gln	Pro	Gly	Lys	Ala	Asp	Leu	Lys	Arg	Glu	Gly
							245			250			255		
Arg	Pro	Leu	Pro	Glu	Gly	Gly	Arg	Gln	Pro	Pro	Ile	Asp	Phe	Arg	Asp
							260			265			270		
Val	Asp	Ile	Gly	Glu	Leu	Ser	Ser	Asp	Val	Ile	Ser	Asn	Ile	Glu	Thr
							275			280			285		

Phe Asp Val Asn Glu Phe Asp Gln Tyr Leu Pro Pro Asn Gly His Pro  
290 295 300  
Gly Val Pro Ala Thr His Gly Gln Val Thr Tyr Thr Gly Ser Tyr Gly  
305 310 315 320  
Ile Ser Ser Thr Ala Ala Thr Pro Ala Ser Ala Gly His Val Trp Met  
325 330 335  
Ser Lys Gln Gln Ala Pro Pro Pro Pro Gln Gln Pro Pro Gln Ala  
340 345 350  
Pro Pro Ala Pro Gln Ala Pro Pro Gln Pro Gln Ala Ala Pro Pro Gln  
355 360 365  
Gln Pro Ala Ala Pro Pro Gln Gln Pro Gln Ala His Thr Leu Thr Thr  
370 375 380  
Leu Ser Ser Glu Pro Gly Gln Ser Gln Arg Thr His Ile Lys Thr Glu  
385 390 395 400  
Gln Leu Ser Pro Ser His Tyr Ser Glu Gln Gln Gln His Ser Pro Gln  
405 410 415  
Gln Ile Ala Tyr Ser Pro Phe Asn Leu Pro His Tyr Ser Pro Ser Tyr  
420 425 430  
Pro Pro Ile Thr Arg Ser Gln Tyr Asp Tyr Thr Asp His Gln Asn Ser  
435 440 445  
Ser Ser Tyr Tyr Ser His Ala Ala Gly Gln Gly Thr Gly Leu Tyr Ser  
450 455 460  
Thr Phe Thr Tyr Met Asn Pro Ala Gln Arg Pro Met Tyr Thr Pro Ile  
465 470 475 480  
Ala Asp Thr Ser Gly Val Pro Ser Ile Pro Gln Thr His Ser Pro Gln  
485 490 495  
His Trp Glu Gln Pro Val Tyr Thr Gln Leu Thr Arg Pro  
500 505